NPS ARCHIVE
1965
REILLY, J.

A DYNAMIC INVENTORY MODEL USING EXPONENTIAL SMOOTHING

JOSEPH V. REILLY

U. S. NAVA[illegible] [illegible] SCHOOL
MO[illegible] [illegible]A

A DYNAMIC INVENTORY MODEL

USING EXPONENTIAL SMOOTHING

by

Joseph V. Reilly, Jr.

Lieutenant, United States Navy

Submitted in partial fulfillment of
the requirements for the degree of

MASTER OF SCIENCE
IN
OPERATIONS RESEARCH

United States Naval Postgraduate School
Monterey, California

1 9 6 5

NPS ARCHIVE
1968
REILLY, T

A DYNAMIC INVENTORY MODEL

USING EXPONENTIAL SMOOTHING

by

Joseph V. Reilly, Jr.

This work is accepted as fulfilling

the thesis requirements for the degree of

MASTER OF SCIENCE

IN

OPERATIONS RESEARCH

from the

United States Naval Postgraduate School

ABSTRACT

Inventory systems may be analyzed using techniques which were originally devised to study the dynamic response of servo-mechanisms. This paper uses these methods to study an inventory system using exponential smoothing to predict a demand which is assumed to have a linear trend with uncorrelated noise components. The statistical properties of the estimator are investigated, and both exact and limiting expressions for its mean and variance are derived. The model is assumed to have a fixed time lag between the placing of a replenishment order and its receipt, and the system response is determined for impulse, step and ramp inputs. The choice of a smoothing constant is discussed, and the effect of the choice on the variance of the forecast equation, and the response time of the system is analyzed.

TABLE OF CONTENTS

TABLE OF SYMBOLS

$I(t)$	Inventory at time t: $t = 1,2, \ldots.$
$X(t)$	Demand during the t th time period
$\hat{S}_t(X;\tau)$	Forecasted demand during the $(t+\tau)$th period
$O(t)$	Replenishment order placed at time t
$R(t)$	Receipt at time t
I_0	Initial Inventory
α	Exponential smoothing constant
$S_t(x)$	Single exponential smoothing operator
$S_t^{[2]}(x)$	Double exponential smoothing operator

1. Introduction.

This paper is a study of a dynamic inventory system using techniques of linear systems analysis(1). The model is an inventory control system with a fixed order point, and a known fixed time lag between the placing of a replenishment order and its receipt. The demand is assumed to be random, with a linear trend and uncorrelated noise components. Exponential smoothing is used to forecast future demand (2). The reorder rule is designed to minimize inventory variation for any sequence of forecasting errors (4). The controlled quantity (inventory) may be either positive or negative. A negative inventory represents a situation where the demand has exceeded the material on hand, and back orders have been placed against future deliveries. In order to facilitate the mathematical manipulations, the system equations will be transformed using generating functions, and the response of the system to deterministic inputs (impulse, step, and ramp) will be investigated. In this paper the term estimate is synonymous with predict.

2. Generating Functions

In order to facilitate analysis of the system, the variables shall be treated as infinite discrete time series, and their generating functions will be obtained. For example if X is the input (demand), with components X(n); n=1,2,..., then the generating function for X is given by,

(2-1)
$$X^T(z) = \sum_{m=0}^{\infty} X(m)\, z^m$$

An important characteristic of generating functions is the following:

If

$$g(n) = \sum_{k=0}^{n} f(k)\,h(n-k) \tag{2-2}$$

where g, f and h are arbitrary discrete time series with components g(n), f(n) and h(n) then

$$\sum_{n=0}^{\infty} g(n) z^n = \sum_{n=0}^{\infty} z^n \sum_{k=0}^{n} f(k)\,h(n-k) \tag{2-3}$$

$$= \sum_{k=0}^{\infty} f(k)\, z^k \sum_{j=0}^{\infty} h(j)\, z^j$$

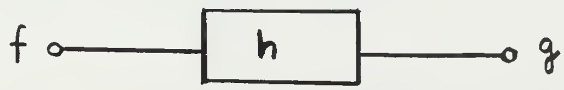
or

$$g^T(z) = f^T(z)\, h^T(z)$$

Thus, the convolution operation on the original variables is analagous to multiplication of the transformed variables.

3. Flow Graphs

Flow graphs are compact "visual devices" that help describe relationships in a linear system where the basic operation performed is convolution. For example equation (2-3) may be represented graphically as:



[illegible]

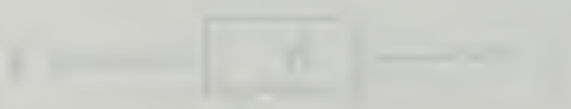
$$[illegible]$$

[illegible]

$$[illegible]$$

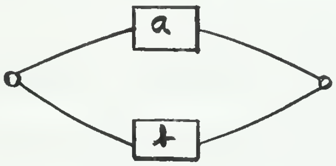
[illegible]

[illegible]

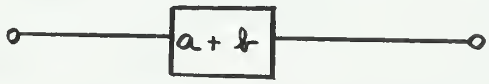


[illegible]

where the input to the system is some descrete time series f, and the output is another discrete time series g which is produced by the convolution of f with the system "transfer function" h. The circles at the input and the output of the system are called "nodes" and any path connecting two nodes is called a "branch". The purpose of introducing this notation is that more complex systems may be simplified into equivalent systems such as pictured above by applying some general equivalence rules. For example

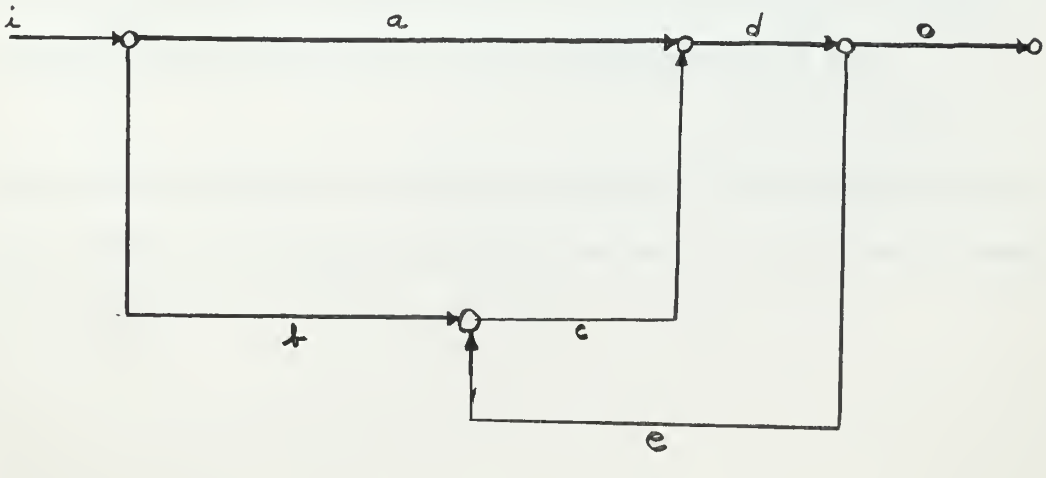


is equivalent to



because of linearity.

A network which is similar to the inventory problem is:



where i is the input to the network and o is the output. The letters above the branches represent the transfer function of their respective branches.

This is similar to networks found in electrical control systems and methods devised for analysis of these physical systems work equally well in the analysis of the inventory system. A first step in the analysis is to find any "loops" in the systems, i.e., a closed path through the network in the direction of the arrows which touches no node more than once in one traversal of the "loop". In the above example cde is a loop. The "loop product" is defined as the negative product of the transfer functions of the branches making up the loop. In the above example -cde is the "loop product". The network determinant (Δ) is defined as 1 plus the sum of all the loop products in the network. In this case:

$$\Delta = 1 - c\ d\ e$$

The final quantities required are the path transmission P_i and the path determinant Δ_i . P_i is simply the product of the branch transfer functions for any simple path between the input node and the output node in the direction of the branch arrows. In the example above there are two path transmissions $P_1 = a\ d$

$$P_2 = b\ c\ d$$

The path determinant is 1 plus the product of all loops in the network which do not share any node in the path. In the above example $\Delta_1 = \Delta_2 = 1$

The network transfer function is then defined as

(3-1) $$T = \frac{\sum_{k} P_k \Delta_k}{\Delta}$$

In the above example

(3-2) $$T_{io} = \frac{ad + bcd}{1 - cde}$$

This result will be used in the analysis of the inventory problem.

4. Exponential Smoothing

The operation of exponential smoothing is a weighted average of past observations used to predict what some future value of the variable will be. The weight given to previous observations decreases geometrically with age. The forecast equation used in this paper is a linear combination of single and double exponential smoothing. It will be shown that a linear combination of single and double smoothing yields an asymtotically unbiased estimate of the demand when the demand is assumed to have a linear trend.

The Single Exponential Smoothing operator is given by:

(4-1) $$\begin{aligned} S_t(X) &= \alpha X(t) + \beta S_{t-1}(X) \\ &= \alpha X(t) + \beta [\alpha X(t-1) + \beta S_{t-2}(X)] \\ &= \cdots \\ &= \alpha \sum_{K=0}^{t-1} \beta^K X(t-K) + \beta^t X_0 \end{aligned}$$

Where X_0 is the initial estimate of the demand and $\beta = 1 - \alpha$

Suppose demand has no trend component, i.e.:

$$X(t) = a + \epsilon_t \qquad t > 0$$

Where a is a constant and ϵ_t is random with

$$E[\epsilon_t] = 0 \qquad \forall t$$

$$E[\epsilon_t^2] = \sigma_\epsilon^2 \qquad \forall t$$

and

$$E[\epsilon_i \epsilon_j] = 0 \qquad \forall i \neq j$$

Then,

$$\text{(4-2)} \quad E[S_t(x)] = E\left[\alpha \sum_{K=0}^{t-1} \beta^K X(t-K) + \beta^t X_0\right]$$

$$= a\alpha \sum_{K=0}^{t-1} \beta^K + \beta^t E[X_0]$$

$$= a - a\beta^t + \beta^t E[X_0]$$

If the initial estimate of the demand is an unbiased estimate based upon past observations then

$E[S_t(x)] = a \quad \forall t$ and, in any event,

$$\lim_{t \to \infty} E[S_t(x)] = a$$

Since smoothing is a linear operation:

(4-3)
$$S_t(x) = S_t(a) + S_t(\epsilon)$$
$$= \alpha a \sum_{K=0}^{t-1} \beta^K + \beta^t a + S_t(\epsilon)$$
$$= a + S_t(\epsilon)$$

(4-4)
$$S_t^2(x) = a^2 + 2a\, S_t(\epsilon) + S_t^2(\epsilon)$$
$$E[S_t^2[x]] = a^2 + E\Big[\alpha^2 \sum_{K=0}^{t-1} \beta^{2K} \epsilon_{t-K}^2 + \alpha^2 \sum_{i=1}^{t-1}\sum_{j=0}^{t-2} \beta^{i+j} \epsilon_{t-i}\epsilon_{t-j} + 2\beta^t \epsilon_0 \sum_{K=0}^{t-1} \beta^K \epsilon_{t-K} + \beta^{2t}\epsilon_0^2\Big]$$
$$= a^2 + \alpha \sigma_\epsilon^2 \left[\frac{1-\beta^{2t}}{1+\beta}\right] + \beta^{2t} \sigma_{\epsilon_0}^2$$

Where $\sigma_{\epsilon_0}^2$ is the variance of the initial estimate of demand.

Hence,
$$Var[S_t(x)] = E[S_t^2(x)] - [E(S_t(x))]^2$$

(4-5)
$$= \alpha \sigma_\epsilon^2 \left[\frac{1-\beta^{2t}}{1+\beta}\right] + \beta^{2t} \sigma_{\epsilon_0}^2$$

It follows that:

$$\lim_{t\to\infty} Var[S_t(x)] = \frac{\alpha}{1+\beta} \sigma_\epsilon^2$$

The above results indicate that exponential smoothing gives an asymtotically unbiased estimate of the demand and, if the criterion for choosing the smoothing constant were minimum variance, the choice $\alpha = 0$ would be optimal. This of course would be the same as using the initial estimate of demand for all future estimates and would negate the most desirable trait of exponential smoothing, i.e.: adjusting the estimate when the time demand changes. It does however indicate that where there is a high degree of confidence in the initial estimate, and it appears that changes in demand are unlikely, a small value of the smoothing constant is preferable.

5. Double Exponential Smoothing

For inventory situations where the demand is assumed to have the form:

$X(t) = a + \epsilon_t$ as discussed above, the single smoothing operator gives a suitable estimate of future demand. If demand is assumed to have a linear trend component, then it will be shown that both single and double smoothing give biased estimates of the demand even in the limit, but a linear combination of the two operators yields an asymtotically unbiased estimate.

For the linear model:

$X(t) = a + bt + \epsilon_t$ where a and b are constants and,

$E[\epsilon_t] = 0$

$E[\epsilon_t^2] = \sigma_\epsilon^2 \; \forall t$ with

$E[\epsilon_i \epsilon_j] = 0 \quad \forall i \neq j$

Let $\xi_t = a + bt$ so that,

$S_t(X) = S_t(\xi) + S_t(\epsilon)$ by linearity.

Now, $\xi_{t-K} = a + bt - bK = \xi_t - bK$

$$S_t(\xi) = \alpha \sum_{K=0}^{t-1} \beta^K \xi_t - b\alpha \sum_{K=0}^{t-1} K\beta^K + \xi_0 \beta^t \tag{5-1}$$

$$= \alpha \xi_t \left[\frac{1-\beta^t}{1-\beta}\right] - b\alpha\beta \left[\frac{(1-\beta^t) - t\alpha\beta^{t-1}}{\alpha^2}\right] + \xi_0 \beta^t$$

$$= \xi_t [1-\beta^t] - \frac{b\beta}{\alpha}\left[1 - \beta^t - \alpha t \beta^{t-1}\right] + \xi_0 \beta^t$$

$$= \xi_t [1-\beta^t] - \frac{b\beta}{\alpha}\left[1 - \beta^t[1-t] - t\beta^{t-1}\right] + \xi_0 \beta^t$$

Since $E[S_t(\epsilon)] = E\left[\alpha \sum_{K=0}^{t-1} \beta^K \epsilon_{t-K} + \beta^t \epsilon_0\right] = 0$

$$E[S_t(x)] = E[S_t(\xi)] + E[S_t(\epsilon)] = S_t(\xi) \tag{5-2}$$

$$\lim_{t\to\infty}\left[E[S_t(x)] - E[X(t)]\right] = \lim_{t\to\infty}\left[S_t(\xi) - a - bt\right] \tag{5-3}$$

$$= -\frac{b\beta}{\alpha}$$

The above result shows that the single smoothing operator does not give an unbiased estimate of the demand in the linear model even asymtotically. Moreover, after a sufficient time period to allow the exponential terms to become negligible the expected value of the single smoothing estimator may be written as;

$$E[S_t(x)] \doteq a + bt - \frac{b\beta}{\alpha} \tag{5-4}$$

[illegible]

[illegible]

[illegible]

$$[illegible]$$

$$[illegible]$$

$$[illegible]$$

$$[illegible]$$

$$[illegible]$$

$$[illegible]$$

$$[illegible]$$

[illegible]

$$[illegible]$$

Similarly for double smoothing:

(5-5) $$S_t^{[2]}(X) = S_t^{[2]}(\xi) + S_t^{[2]}(\epsilon)$$

(5-6) $$S_t^{[2]}(\xi) = \alpha \sum_{K=0}^{t-1} \beta^K S_{t-K}(\xi) + \beta^t S_0^{[2]}(\xi)$$

$$S_{t-K}(\xi) = [\xi - b k][1 - \beta^{t-k}] - \frac{b\beta}{\alpha}[1 + \beta^{t-k}[t-k-1] - [t-k]\beta^{t-k-1}] + \beta^{t-k}\xi_0$$

$$S_t^{[2]}(\xi) = \alpha \xi_t \sum_{K=0}^{t-1} \beta^k - \alpha b \sum_{K=0}^{t-1} k\beta^k - b\beta \sum_{k=0}^{t-1} \beta^K - b\beta \sum_{K=0}^{t-1} \beta^t [t-k-1] + b\beta \sum_{K=0}^{t-1} \beta^{t-1}[t-K] + \beta^t \xi_0$$

$$= \xi_t [1-\beta^t] - \frac{b\beta}{\alpha}[1 + \beta^t[t-1] - t\beta^{t-1}] - \frac{b\beta}{\alpha}[1-\beta^t] + \frac{t[t+1] b \alpha \beta^t}{2} + \beta^t \xi_0$$

Since $E[S_t^{[2]}(\epsilon)] = 0$

(5-7) $$E[S_t^{[2]}(X)] = S_t^{[2]}(\xi)$$

and $$\lim_{t\to\infty} E[S_t^{[2]}(X) - X(t)] = \lim_{t\to\infty}[S_t^{[2]}(\xi) - \xi_t] = -\frac{2b\beta}{\alpha}$$

Here again the estimate is biased and the limiting mean may be written approximately as:

(5-8) $$E[S_t^{[2]}(X)] \doteq a + bt - \frac{2b\beta}{\alpha}$$

Using these approximations it can be seen that:

$$\text{(5-9)} \qquad \hat{a} = 2\, S_t[X] - S_t^{[2]}[X]$$

$$\text{(5-10)} \qquad \hat{b} = \frac{\alpha}{\beta}\left[S_t[X] - S_t^{[2]}[X] \right]$$

are asymtotically unbiased estimates of the parameters in the linear No model. Although the estimates are not unbiased the bias does decrease exponentially with time. Using these estimators, the forecast of demand at time $(t + \tau)$ becomes:

$$\text{(5-11)} \qquad \hat{S}_t(X;\tau) = \left[2 + \frac{\alpha\tau}{\beta}\right] S_t(X) - \left[1 + \frac{\alpha\tau}{\beta}\right] S_t^{[2]}(X)$$

The generating functions for exponential smoothing will now be introduced. As given in equation (4-1), single smoothing may be written as:

$$S_t(X) = \alpha \sum_{K=0}^{t-1} \beta^K X(t-K) + \beta^t X_0$$

which is the convolution of the single smoothing operator with the discrete time series of demand. In the limiting case, i.e., where the initial demand time becomes negligible (or setting $X_0 = 0$), the transformed equation may be written as:

$$\hat{S}(z) = X^T(z)\, S^T(z)$$

where $S^T(Z)$ is the transform of an exponential decay function, i.e.:

$$S_t = \alpha \beta^t$$

may be transformed to

$$S^T(z) = \alpha \sum_{n=0}^{\infty} \beta^n z^n = \frac{\alpha}{1-\beta z} \tag{5-12}$$

Similarly for second order smoothing:

$$S_t^{[2]}(x) = \alpha \sum_{K=0}^{t-1} \beta^K S_{t-K}(x) + \beta^t S_0^{[2]}(x)$$

may be recognized as the convolution of first order smoothing with itself for large t, and in the transform domain this is the square of the transform for first order smoothing:

$$S^{T[2]}(z) = \frac{\alpha^2}{(1-\beta z)^2} \tag{5-13}$$

It now remains to find the variance of the forecast equation. First the variance of the second order smoothing operator will be found.

For the linear model:

$$S_t^{[2]}(x) = S_t^{[2]}(\xi) + S_t^{[2]}(\epsilon)$$

Hence Var $\left[S_t^{[2]}(x)\right] = E\left[S_t^{[2]}(\epsilon)\right]^2$ since $S_t^{(2)}(\ )$ is constant

and $E\left[S_t^{[2]}(\epsilon)\right] = 0$

Now, $$S_t^{[2]}(\epsilon) = \alpha \sum_{K=0}^{t-1} \beta^K S_{t-K}(\epsilon) + \beta^t \epsilon_0'$$

where $$\epsilon_0' = S_0^{[2]}(\epsilon)$$

and, $$S_{t-K}(\epsilon) = \alpha \sum_{j=0}^{t-K-1} \beta^j \epsilon_{t-K-j} + \beta^{t-K} \epsilon_0$$

So that, $$S_t^{[2]}(\epsilon) = \alpha^2 \sum_{K=0}^{t-1} \beta^K \sum_{j=0}^{t-K-1} \beta^j \epsilon_{t-K-j} + \alpha \sum_{K=0}^{t-1} \beta^t \epsilon_0 + \beta^t \epsilon_0'$$

$$= \alpha^2 \sum_{K=0}^{t-1} \beta^K \sum_{j=0}^{t-K-1} \beta^j \epsilon_{t-K-j} + \alpha t \beta^t \epsilon_0 + \beta^t \epsilon_0'$$

The second moment is:

(5-14) $$E\left[S_t^{[2]}(\epsilon)\right]^2 = E\left[\alpha^2 \sum_{K=0}^{t-1} \beta^K \sum_{j=0}^{t-K-1} \beta^j \epsilon_{t-K-j} + \alpha t \beta^t \epsilon_0 + \beta^t \epsilon_0'\right]^2$$

$$= \alpha^4 \sigma_\epsilon^2 \sum_{K=0}^{t-1} \beta^{2K} \sum_{j=0}^{t-K-1} \beta^{2j} + \alpha^2 t^2 \beta^{2t} \sigma_{\epsilon_0}^2 + \beta^{2t} \sigma_{\epsilon_0'}^2$$

where $\sigma_{\epsilon_0}^2$ is the variance of the initial estimate for first order smoothing and $\sigma_{\epsilon_0'}^2$ is the variance for second order smoothing. The summation in the first term of equation (5-14) may be evaluated directly or it can be recognized that this is the sum of the first t squared terms of the impulse response of the second order smoothing operator, i.e.;

$$S^{T[2]}(z) = \frac{\alpha^2}{(1-\beta z)^2}$$

where $S^{T[2]}(z)$ is the Z transform for the second order smoothing operator. If we denote the impulse response by F(z);

$$F(z) = X^T(z)\, S^{T[2]}(z)$$

where $X^T(Z) = 1$ (the Z transform of a unit impulse at $t = 0$).

The impulse response is:

$$f(t) = \alpha^2 (t+1)\, \beta^t$$

and the sum of the first t squared terms is

$$\sum_{K=0}^{t-1} f(K) = \alpha^4 \sum_{K=0}^{t-1} (K+1)^2 \beta^{2K}$$

$$= \frac{d^2}{d\Gamma^2} \Gamma^2 \sum_{K=0}^{t-1} \Gamma^K - \frac{d}{d\Gamma} \Gamma \sum_{K=0}^{t-1} \Gamma^K$$

where $\Gamma = \beta^2$; therefore:

(5-15)
$$E\left[S_t^{[2]}(\epsilon)\right]^2 = \beta^{2t}\sigma_{\epsilon_0'}^2 + \alpha^2 t^2 \beta^{2t} \sigma_{\epsilon_0}^2 + \alpha^4 \sigma_\epsilon^2 \Bigg[\frac{1-2\beta^2}{[1-\beta^2]^2} + \frac{2[2\beta^2 - \beta^4]}{[1-\beta^2]^3} - \frac{[t+1][t+2]\beta^{2t}}{[1-\beta^2]} - \frac{2\beta^{2(t+1)}[t+2-(t+1)\beta^2]}{[1-\beta^2]^3} + \frac{2\beta^{2(t+1)} + \beta^{2t}}{[1-\beta^2]^2} \Bigg]$$

and in the limit this becomes:

$$\lim_{t \to \infty} E\left[S_t^{[2]}(\epsilon)\right]^2 = \frac{\alpha[1+\beta^2]}{[1+\beta]^3}\, \sigma_\epsilon^2$$

so that

(5-16)
$$\operatorname{Var}\left[S_t^{[2]}(X)\right] = E\left[S_t^{[2]}(\epsilon)\right]^2 \doteq \frac{\alpha[1+\beta^2]}{[1+\beta]^3}\, \sigma_\epsilon^2$$

The transfer function approach may also be used to evaluate the variance of the forecast equation. In this case

$$\hat{S}^T(z) = \left[2 + \frac{\alpha\tau}{\beta}\right]\frac{\alpha}{1-\beta z} - \left[1 + \frac{\alpha\tau}{\beta}\right]\frac{\alpha^2}{(1-\beta z)^2}$$

The impulse response is:

$$\left[2 + \frac{\alpha\tau}{\beta}\right]\alpha\beta^{t} - \left[1 + \frac{\alpha\tau}{\beta}\right][t+1]\,\beta^{t}$$

and the sum of the squares of the response terms is

$$\sum_{K=0}^{t-1}\left[\left[2 + \frac{\alpha\tau}{\beta}\right]\alpha - \alpha^{2}\left[1+\frac{\alpha\tau}{\beta}\right][K+1]\right]^{2}\beta^{2K}$$

The limiting or asymtotic variance of the estimator then becomes

$$\text{(5-17)}\quad \mathrm{Var}\left[\hat{S}_{t}(x)\right] \doteq \frac{\left[1 + 4\beta + 5\beta^{2} + 2\alpha(1+3\beta)\tau + 2\alpha^{2}\tau^{2}\right]\alpha\,\sigma_{\epsilon}^{2}}{\left[1+\beta\right]^{3}}$$

It is then apparent that the proper linear combination of single and double exponential smoothing will give an asymtotically unbiased estimate of demand where there is a linear trend, and the noise components are uncorrelated. As can be seen in the derivation of the moments of the smoothing operator the variance of the operator is a weighted sum of the variance of the observed demand and the variance of the initial estimate. As more observations are obtained the variance then approaches the limiting variance as given by equation (5-17) for the forecast equation of the linear model. In the limiting case it can be seen that small values of α give the smallest variance. It will be seen later in the paper that although small variance is a desirable characteristic of any forecast, the response of the forecast equation to changes in demand pattern is also a function of α and if α is chosen to be small the model will be slow in responding to changes in the demand.

6. Fixed Lag Inventory Model:

The problem to be considered is an inventory system where there are fixed order points. In practice this would correspond to an activity which finds it more economical or necessary to adhere to a regular order schedule, and can only place replenishment orders at the beginning of each input period. Specific characteristics of the system are:

1) The item has high usage, and there are indications of a linear trend in the demand.

2) There is no seasonal variation in demand, and no correlation in the error terms.

3) The order points are fixed and equally spaced.

4) When demand exceeds the inventory on hand, back orders are placed against future deliveries.

5) There is a fixed time lag between placing a replenishment order and receipt of the order. Time lag is measured in demand periods.

6) Time is measured from the end of the demand period.

7) A replenishment order placed at time t_k may be used to fill demand generated during demand period $(t_k + T + 1)$, where the time lag is T.

8) Oversupply may be returned to the vendor; i.e., a negative order quantity may be generated.

9) The forecast equation introduced in section (5) will be used to predict future demand.

10) The forecast at time k will predict demand during period $k + T + 1$.

The following equations describe the model:

(6-1) $$I(t) = I_0 + \sum_{i=1}^{t} R_i - \sum_{j=1}^{t} X(j)$$

(6-2) $$\hat{S}_t(x;\tau) = \left[2 + \frac{\alpha\tau}{\beta}\right] S_t(x) - \left[1 + \frac{\alpha\tau}{\beta}\right] S_t^{[2]}(x)$$

where:

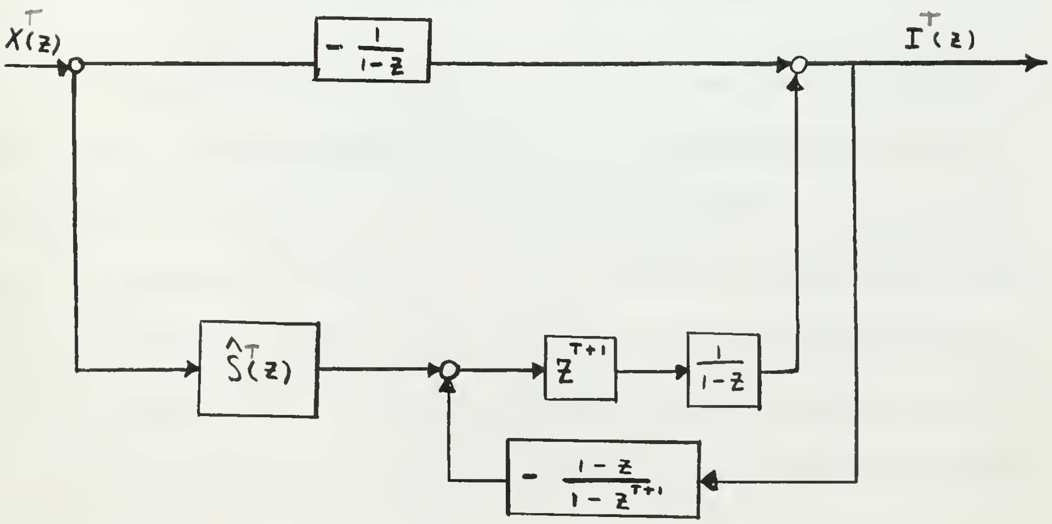
(6-3) $$S_t(x) = \alpha X(t) + \beta S_{t-1}(x)$$

(6-4) $$S_t^{[2]}(x) = \alpha S_t(x) + \beta S_{t-1}^{[2]}(x)$$

(6-5) $$O(t) = \sum_{j=0}^{T} \hat{S}_{t-j+(T+1)}(x) - \sum_{j=1}^{T} O(t-j) - I(t)$$

(6-6) $$R(t) = O(t-(T+1))$$

The flow graph for this model is:



where:

$$\hat{S}^{T}(z) = \left[2 + \frac{\alpha[T+1]}{\beta}\right]\frac{\alpha}{1-\beta z} - \left[1 + \frac{\alpha[T+1]}{\beta}\right]\frac{\alpha^2}{(1-\beta z)^2} \tag{6-7}$$

$$= \frac{\alpha\left[(2+\alpha T) - (2+\alpha(T-1))z\right]}{(1-\beta z)^2}$$

which is a linear combination of the transforms for single and double exponential smoothing. The network determinant is:

$$\Delta = 1 + \frac{z^{T+1}}{(1-z)(1-z^{T+1})} = \frac{1}{1-z^{T+1}}$$

The path transmissions and path determinants are:

$$P_1 = -\frac{1}{1-z} \qquad \Delta_1 = 1$$

$$P_2 = \frac{\hat{S}(z)\, z^{T+1}}{1-z} \qquad \Delta_2 = 1$$

and the system transfer function is

$$S = \frac{P_1\Delta_1 + P_2\Delta_2}{\Delta} = \frac{\alpha z^{T+1}(1-z^{T+1})\left[(2+\alpha T) - (2+\alpha(T-1))z\right]}{(1-z)(1-\beta z)^2} - \frac{1-z^{T+1}}{1-z} \tag{6-8}$$

The purpose of describing the inventory model in terms of flow graphs and generating functions is to determine the dynamic response of the system to possible demand patterns. It was seen that in examining the moments of the forecast equation the variance of the forecast equation was a function of the smoothing constant α and the lead time T. On examining the dynamic response of the system to deterministic inputs it will be seen that the same para-

[illegible]

meters determine the response time (time until all back orders are filled). The model will be tested with three deterministic inputs (impulse, step, ramp) and the model will be checked for response to these inputs, as well as the steady state error.

In order to facilitate the determination of steady state error the following theorem from (3) is introduced.

Let F Be the transform of f and suppose

$F(Z)$ converges as $Z \to 1$ where Z is real and

$|Z| < 1$ and $f(n) = 0$ for $n < 0$. Then:

$$\lim_{n\to\infty} f(n) = f(\infty) = \lim_{Z\to 1} (1 - Z)\, F(z)$$

6.1 Impulse Response

The impulse response for the inventory model in transform notation becomes

$$I^T(z) = S\, X^T(z) \qquad\qquad X^T(z) \equiv 1$$

Hence by the above theorem;

$$I(\infty) = \lim_{z\to 1} (1-z)\left[\frac{(1-z^{T+1})\,\alpha\, z^{T+1}\,[(2+\alpha T) - (2+\alpha[T-1])\, z}{(1-z)(1-\beta z)^2} - \frac{1-z^{T+1}}{1-z}\right] = 0$$

Here $I(\infty)$ denotes the "steady state" inventory.

In order to obtain the inverse transform, the transfer function may be expanded into a sum of partial fractions:

$$I^T(z) = z^{T+1}(1-z^{T+1})\left[\frac{-\beta}{(1-\beta z)} + \frac{\alpha[1+\alpha T]}{(1-\beta z)^2} + \frac{1}{1-z}\right] - \frac{1-z^{T+1}}{1-z}$$

Then solving for the inverse:

$$I(t) = -\beta\left[\beta^{t-[T+1]}\mu(t-(T+1)) - \beta^{t-2[T+1]}\mu(t-2(T+1))\right]$$

(6-9) $$+\ \frac{\alpha[1+\alpha T]}{\beta}\left[(t-T)\beta^{t-T}\mu(t-T) - (t-(2T+1))\beta^{t-(2T+1)}\mu(t-2T+1)\right]$$

$$+\ 2\mu(t-(T+1)) - \mu(t-2(T+1)) - \mu(t)$$

where $\mu(K) = 1 \qquad K \geq 0$

$\qquad\qquad = 0 \qquad K < 0$

In addition to the fact that there is no steady state error, the expression for the impulse response also shows that if the system was receiving a constant demand for a long period of time and there was a one time increase in that demand the system would respond in the minimum amount of time since:

$$I(t) = -1 \quad ; \ t < T+1$$

$$I(T+1) = \alpha(2+\alpha T)$$

In other words, there would be a back order for a period equal to the delivery time but all demand would be met by time T + 1, and then in fact there would be an excess of material $(\alpha(2+\alpha T))$

6.2 Response To A Step Input $\left(X^T(z) = \frac{1}{1-z}\right)$

Testing the system with a step input gives an indication of system performance when there is a sudden increase in demand and demand continues at this new higher level. The results which are derived in this example apply equally to the case where there is a sudden drop in the demand level.

$$I^{T}(z) = S\,X^{T}(z) = \frac{S}{1-z}$$

$$I(\infty) = \lim_{z \to 1} (1-z)\frac{S}{1-z} = 0$$

Using the same partial fraction expansion as above the inverse transform yields:

(6-10)

$$I(t) = -\frac{\beta}{\alpha}\Big[(1-\beta^{t-(T+1)})\,u(t-(T+1)) - (1-\beta^{t-2(T+1)})\,u(t-2(T+1))$$

$$+\frac{[1+\alpha T]}{\alpha}\Big[1-[1+(t-T)\alpha]\beta^{t-T}u(t-(T+1)) - [1 -$$

$$[1+(t+1-2(T+1)\alpha]\beta^{t-2T-1}u(t-2(T+1))\Big]$$

$$+2(t-T)u(t-(T+1)) - (t-(2T+1))u(t-2(T+1))$$

$$-(t+1)u(t)$$

As expected there is again negative inventory after the initial change in demand. At the start:

$$I(t) = -(t+1)\ ; \quad 0 \le t < T+1$$

The first order is received at (T + 1) in the amount $2 + \alpha(1 + \alpha T)$. It appears then that if the criterion for selecting a smoothing constant was minimum response time, $\alpha = 1$ would be the optimal choice. This would be the extreme case where no weight was given to past demand and the forecast was based upon only the last recorded demand. This of course would be impractical since actual demand will normally have a noise component (i.e. be probabalistic rather than deterministic) and as shown previously a choice of $\alpha = 1$ would make the variance of the forecast equation maximum.

It is then evident that no single value of the smoothing constant will satisfy both the requirement of minimum forecast variance and minimum response time.

6.3 Response To A Ramp

This is the type input for which the forecast equation was designed. The reasons for looking at the step and impulse responses were to see how the system performed under other than designed conditions. Since the system is linear and time invariant the performance of the system to any combination of inputs may be analyzed by looking at each input separately and obtaining the actual response by adding the system response to each of the components of the input.

For the ramp input: $X^T(z) = \frac{z}{(1-z)^2}$

$$I^T(z) = S X^T(z) = \frac{S z}{(1-z)^2}$$

$$I(\infty) = \lim_{z \to 1} (1-z) \frac{S z}{(1-z)^2} = 0$$

Expanding $I^T(z)$ as before and solving for $I(t)$

$$\begin{aligned} I(t) = {} & -\frac{\beta}{\alpha}\left[t-(T+1)-\frac{\beta}{\alpha}\left(1-\beta^{t-(T+1)}\right)\mu(t-(T+1))\right] \\ & + \frac{\beta}{\alpha}\left[t-2(T+1)-\frac{\beta}{\alpha}\left(1-\beta^{t-2(T+1)}\right)\mu(t-2(T+1))\right] \\ & + \frac{[1+\alpha T]}{\alpha}\left[t-(T+1)-\frac{2\beta}{\alpha}+\left(\frac{2}{\alpha}+t-(T+1)\right)\beta^{t-T}\right] \\ & \mu(t-(T+1)) - \frac{[1+\alpha T]}{\alpha}\left[t-(T+1)-\frac{2\beta}{\alpha}+\left(\frac{2}{\alpha}+t-2(T+1)\right)\beta^{t-2T-1}\right] \\ & \mu(t-2(T+1)) + \sum_{j=T+1}^{2T+1}(t-j)\,\mu(t-j) - \sum_{K=0}^{T}(t-K)\,\mu(t-K) \end{aligned} \tag{6-11}$$

It is difficult to analyze response time from the above expression but if a ramp input is looked upon as a sequence of impulse inputs where the magnitude of the impulse at time t is $(t + 1)$ the superposition property of a linear time invariant system leads to the conclusion that the larger the value of α, the smaller the response time.

6.4 Order Rule

As shown by Vassian (3), the order rule utilized in the present model gives minimum inventory variation for any sequence of forecasting errors.

7. Distributed Lag Inventory Model

The preceding model with a fixed lag time was characterized by a system response which was a function of the smoothing constant and an inventory which tended to overshoot the demand when α was chosen to be large in order to decrease the response time. The same problem will now be examined when the lag time is distributed rather than fixed (5). In the context of the inventory model we now have that;

(7-1) $$R(t) = \sum_{j=1}^{t} O(j)\, f(t-j)$$

where; $$f(t\,;p) = p\, q^{t}$$

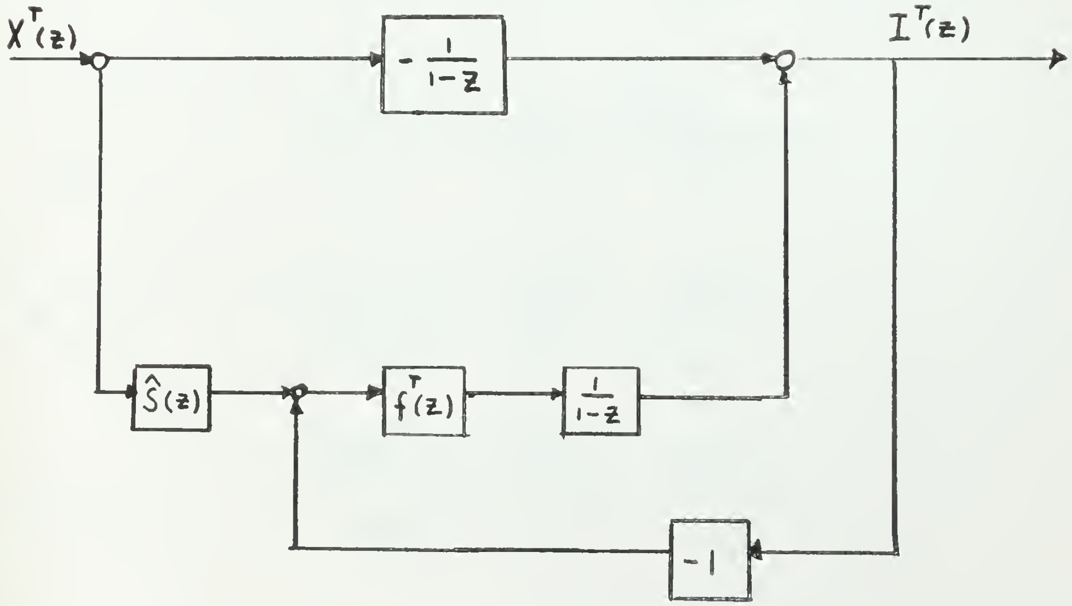
is the probability that an order is received exactly t days after it is placed.

The order rule will now be;

$$(7\text{-}2) \qquad O(t) = \hat{S}_t(X;\tau) - I(t)$$

and the forecast equation will predict the demand at $\tau = q/p$ (the expected delivery time).

The flow graph is:



and

$$f^T(z) = \frac{p}{1-qz}$$

This network has the transfer function:

$$S = \frac{1}{g} \, \frac{[1-z]^2 \, [\beta^2 g z - (\beta^2 + 2\beta g - 2\beta^2 g)]}{[1-\beta z]^2 \, [z^2 - \frac{g+1}{g} z + \frac{p+1}{g}]} \tag{7-3}$$

$$= \frac{A_1}{(z - 1/\beta)} + \frac{A_2}{(z - 1/\beta)^2} + \frac{A_3}{(z - r_1)} + \frac{A_4}{(z - r_2)}$$

where r_1 and r_2 are the roots of the second degree polynominal in the denominator of S, and;

$$A_1 = \frac{d}{dz}\left[\frac{[1-z]^2 \, [z - (1/g + 2\alpha/\beta)]}{z^2 - \frac{g+1}{g} z + \frac{p+1}{g}} \right]_{z = 1/\beta}$$

$$A_2 = \frac{[\alpha/\beta]^2 \, [1/\beta - 1/g - \frac{2\alpha}{\beta}]}{[1/\beta^2 - g+1/g\beta + \frac{p+1}{g}]}$$

$$A_3 = \frac{\left[\frac{g-1-R}{2g}\right] \left[\frac{g-1+R}{2g} - \frac{2\alpha}{\beta}\right]}{\left[\frac{g+1+R}{2g} - 1/\beta\right]^2 \, [R/g]}$$

$$A_4 = \frac{\left[\frac{g-1+R}{2g}\right] \left[\frac{g-1-R}{2g} - \frac{2\alpha}{\beta}\right]}{\left[\frac{g+1-R}{2g} - 1/\beta\right]^2 \, [-R/g]}$$

,where $R = \sqrt{5g^2 - 6g + 1}$

[illegible]

$$[illegible]$$

[illegible]

[illegible]

[illegible]

$$[illegible]$$

$$[illegible]$$

$$[illegible]$$

$$[illegible]$$

$$[illegible]$$

7.1 Impulse Response

$$X^T(z) \equiv 1$$

$$I^T(z) = S$$

(7-4) $$I(t) = -\frac{A_1}{\beta}\beta^t + \frac{A_2}{\beta^2}(t+1)\beta^t - A_3 r_1^{-(t+1)} - A_4 r_2^{-(t+1)}$$

$$I(\infty) = 0$$

Just as with the fixed time lag model there is no steady state error, but the transient response does exhibit a property not evident in the fixed lag model. For $0 \leq q \leq .2$ R is real and all of the terms in $I(t)$ are exponential decay. For $.2 < q < 1.0$ R is imaginary and the coefficients A_3 and A_4 are related by:

$$A_3 = \tilde{A}_4$$

and $$r_1 = \tilde{r}_2$$

so that when written in exponential form:

$$A_3 = Ce^{j\theta} \qquad A_4 = Ce^{-j\theta}$$

$$r_1 = De^{j\varphi} \qquad r_2 = De^{-j\varphi}$$

and the last two terms in $I(t)$ become

$$I(t) = \ldots\ldots - CD^{-(t+1)}\left[e^{j[\theta+(t+1)\varphi]} + e^{-j[\theta+(t+1)\varphi]}\right]$$

$$= \ldots\ldots - CD^{-(t+1)}\cos\left[\theta + (t+1)\varphi\right]$$

The response then is the sum of terms containing exponential decay and damped oscillation.

The results for a step and ramp input as seen below, also contain terms which are damped oscillations for the values of q given above. This result is not unexpected. Damped oscillation is present in physical systems in which there is delay in response. The lag in receipt in the inventory system is analagous to inertia in a mechanical system or inductance in an electrical system, and both of these physical systems will have transient responses containing damped oscillation under certain conditions which are analagous to the relationship of q in the inventory system.

The steady state error for both the step and ramp input is zero in the distributed lag system and, other than this tendency to oscillate, the results are similar to those in the fixed lag model.

7.2 Inventory Function For A Step Input (Distributed lag); when demand is a step the response is;

$$I(t) = \frac{A_1}{\alpha\beta}\left[1-\beta^{t+1}\right] + \frac{A_2}{\alpha^2\beta^2}\left[1-(1+(t+1)\alpha]\,\beta^{t+1}\right] + \frac{A_3}{1-r_1}\left[1-r_1^{-(t+1)}\right] + \frac{A_4}{1-r_2}\left[1-r_2^{-(t+1)}\right] \tag{7-5}$$

7.3 Inventory Function For A Ramp Input (Distributed lag); when demand is a ramp the response is;

$$I(t) = -\frac{A_1}{\beta}\left[t - \frac{\beta}{\alpha}\left[1-\beta^{t}\right]\right] + \frac{A_2}{\beta^2}\left[t - \frac{2\beta}{\alpha} + \left[\frac{2}{\alpha} + t\right]\beta^{t+1}\right] \tag{7-6}$$

$$+ A_3\left[\frac{r_1}{(1-r_1)^2} + \frac{t+1}{(1-r_1)} - \frac{1}{(1-r_1)^2}\, r_1^{-t}\right]$$

$$+ A_4\left[\frac{r_2}{(1-r_2)^2} + \frac{t+1}{(1-r_2)} - \frac{1}{(1-r_2)^2}\, r_2^{-t}\right]$$

7.4 Summary

The distributed lag model is probably a more realistic description of a practical inventory problem in that the delivery time for replenishment orders would seldom be a fixed constant as was assumed in the first model. Practical inventory problems are characterized by the damped oscillations of the distributed lag model. When the delivery time is not fixed, it should be expected that there will be periods of shortage and inventory excess caused by variations in delivery time. As can be seen in the above equations response time is now a function of the smoothing constant α and β, the parameter of the geometric distribution of delivery time. The inventory equations could be analyzed directly in order to choose an acceptable value for the smoothing constant, but as long as it is recognized that there will be some oscillation no matter how the smoothing constant is chosen it does not appear to be invalid to use the fixed lag formulation of the problem, where the fixed lag is chosen to be the expected value of the random variable (lag time).

8. Selection of a Smoothing Constant

In choosing a smoothing constant for the model it must be recognized that no single value of α is optimal for all demand patterns. The impulse response for the fixed lag model shows the relationship which exists between α and the time lag. It will be recalled that for this model:

$$I(t) = -1 \qquad 0 \leq t < T+1$$

$$I(T+1) = \qquad \alpha(2 + \alpha T)$$

This indicates that when there is noise present in the demand a long time lag will tend to amplify the fluctuations in inventory caused by the noise. This can be offset somewhat by choosing a smaller α when T is increased, but when α is reduced the response time is increased. The effects of decreasing α however would be felt in the sensitivity of the forecast to changes in parameter. If the slope of the trend component were to increase, a smaller value of α would cause the forecast to lag behind the demand for a longer period of time. Any final choice of α would also have to take into consideration the "cost" of stock outages vs the holding cost of excess inventory. The model considered controlled the inventory about the zero level and in most practical cases a non zero safety level of stock would be maintained. Any non zero safety level would alleviate the response time problem since this buffer stock would compensate for forecast errors during a period of change in demand pattern. All of these factors seem to favor

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

a small smoothing constant to increase system stability, and a safety stock level to compensate for the increased response time.

9. Conclusions

The use of flow graphs and generating functions provide a simple method of obtaining some insight into the characteristics of an actual inventory system. Testing the model with deterministic inputs allows the inventory manager to check his reorder rule for stability and to decide if his forecast equation is satisfactory as regards response to fluctuations in demand patterns. The primary advantage of using exponential smoothing to predict demand is its flexibility and its economy of computer storage space where it is used in a mechanized activity. The flexibility allows the inventory manager to make trade offs between response time and forecast variance by simply adjusting the smoothing constant. If the demand pattern changes at any time, he may increase or decrease α , and if this a temporary change in pattern he may at any time return to the former value of α . The economy of storage space is apparent when exponential smoothing is compared to the use of a moving average to predict demand. Where the moving average estimator using the last N observations of demand would require N storage locations for each item of inventory, exponential smoothing requires only one location to retain the last smoothed value. In this regard, if a moving average estimator provides satisfactory predictions for a given inventory situation, an exponential smoothing

estimator can easily be constructed which will have the same limiting variance as the moving average estimator.

For example if a moving average has been used to predict demand for a constant model; i.e.:

$$X(t) = a + \epsilon_t$$

where the noise is uncorrelated with variance σ_ϵ^2. The moving average estimator is

$$\bar{X}_t = \frac{1}{N} \sum_{j=t-N+1}^{t} X_j \tag{8-1}$$

and

$$Var[\bar{X}_t] = \frac{\sigma_\epsilon^2}{N} \tag{8-2}$$

as shown in section (4) the limiting variance for single smoothing in the case of the constant model was:

$$Var\left[S_t(x)\right] \doteq \frac{\alpha}{1+\beta}\sigma_\epsilon^2$$

so that if α is chosen so that

$$\alpha = 2/N+1 \tag{8-3}$$

the limiting variances of the two estimators are identical. A similar criterion could be established for the linear model and α could be chosen to make the variance of the estimate approximately the same as that for an acceptable moving average operator.

BIBLIOGRAPHY

1. Howard, R. A. Systems Analysis of Linear Systems; Multistage Inventory Models and Techniques. Stanford University Press, 1963.

2. Brown, R. G. Smoothing, Forecasting and Prediction of Discrete Time Systems. Prentice Hall, 1963

3. Brown, W. M. Analysis of Linear Time Invariant Systems McGraw-Hill, 1963

4. Vassian, H. J. Application of Discrete Variable Servo Theory to Inventory Control. Journal of the Operations Research Society of America v. 3, August 1955: 272-282.

5. Simon, H. A. On the Application of Servomechanism Theory in the Study of Production Control. Econometrica v. 20, April 1952: 247-268.

thesR3294
A dynamic inventory model using exponent
3 2768 001 01310 5
DUDLEY KNOX LIBRARY